Received: 24 August 2016 | Revised: 23 January 2017 | Accepted: 30 March 2017

DOI: 10.1111/phc3.12423

WILEY

ARTICLE

Multisensory processing and perceptual consciousness: Part II

Robert Eamon Briscoe

Ohio University

Correspondence
Briscoe, Robert, Philosophy, Ohio University
Email: rbriscoe@gmail.com

Abstract

The first part of this survey article offered a cartography of some of the more extensively studied forms of multisensory processing. In this second part, I turn to examining some of the different possible ways in which the structure of conscious perceptual experience might also be characterized as multisensory. In addition, I discuss the significance of research on multisensory processing and multisensory consciousness for philosophical debates concerning the modularity of perception, cognitive penetration, and the individuation of the senses.

1 | INTRODUCTION

Philosophical and psychological research on perception has historically tended to proceed on a sense-by-sense basis, treating the perceptual modalities as functionally and anatomically independent channels of information about the world (but for an early exception, see Stratton, 1897, 1899). This picture of perception now looks to be substantially incorrect. Numerous experimental studies as well as recently influential causal inference models of perception provide reason to think that multisensory processing in the brain is the norm rather than the exception (Ernst, 2012; Kayser & Shams, 2015; Körding et al., 2007; Rohde, van Dam, & Ernst, 2016; Shams, 2012); that the senses adaptively interact with one another at both early and late stages of perceptual processing (Ghazanfar & Schroeder, 2006; Kupers, Pietrini, Ricciardi, & Ptito, 2011; Shams & Kim, 2010); and that integrating, comparing, or otherwise combining sources of information from different modalities serves both to optimize the estimation of environmental attributes as well as the control of bodily actions (for reviews, see Stein & Meredith, 1993; Calvert, Spence, & Stein, 2004; Spence & Driver, 2004; Trommershäuser, Kording, & Landy, 2011; and Stein, 2012).

It is important, however, to distinguish between multisensory processing and its possible effects on perceptual consciousness (Deroy, Chen, & Spence, 2014; Macpherson, 2011c; Mudrik, Faivre, & Koch, 2014; O'Callaghan 2008, 2012, forthcoming). That the senses pervasively interact with one another at the subpersonal, information-processing level does not by itself entails that experiences produced by such interaction are multisensory in theoretically interesting ways. Interactions between the senses might always result in perceptual experiences that are unimodal both in respect of their phenomenal character and representational content.

The first part of this survey article presented a cartography of some of the more extensively studied forms of multisensory processing. In this second part, I turn to examining some of the different possible ways in which the structure of conscious perceptual experience might also be characterized as multisensory. In addition, I discuss the

 © 2017 The Author(s) Philosophy Compass © 2017 John Wiley & Sons Ltd

significance of research on multisensory processing and multisensory consciousness for philosophical debates concerning the modularity of perception, cognitive penetration, and the individuation of the senses.

2 | IS ALL PERCEPTUAL EXPERIENCE MODALITY SPECIFIC?

Casey O'Callaghan refers to the following traditional view in the philosophy of perception as the "composite snapshot conception":

> *...one's total perceptual experience at a time is an assemblage or composite of modality-specific experiences. Perceptual experience comprises discrete, modality-specific components or 'snapshots'. Each such modality-specific experience has its own recognizable and distinctive character (O'Callaghan, 2008: 321).*

For present purposes, I shall interpret the composite snapshot conception as comprising two main claims. First, the phenomenal character of a subject's overall perceptual experience at a time is always "exhausted by that which could be instantiated by a corresponding merely visual, merely auditory, merely tactual, merely gustatory, or merely olfactory experience, plus whatever accrues thanks to simple co-consciousness" (O'Callaghan, 2015: 562).[1] Call this claim the "composition thesis." Second, all perceptual phenomenal character is modality specific. Call this claim the "distinctiveness thesis." The distinctiveness thesis can be formulated in stronger and weaker terms. According to what O'Callaghan, 2015 refers to as "local" distinctiveness, every phenomenal feature is instantiated by perceptual experiences of just one modality. No aspect of visual phenomenal character, for example, could be instantiated by an experience of touch or audition. A proponent of what O'Callaghan calls "regional" distinctiveness, in contrast, is committed to the weaker claim that within each modality, there are certain phenomenal features that are distinctive to that modality. On this view, it is only the overall phenomenal character of a perceptual experience of a given modality that is distinctive to that modality (O'Callaghan, 2015: 558).

The composite snapshot conception of perceptual consciousness has been challenged with a wide range of empirically and phenomenologically motivated objections. In the remainder of this section, I critically examine these objections in turn, beginning with the argument from common sensibles.

2.1 | The argument from common sensibles

Common sensibles are properties and relations that can be perceived through more than one modality. 3-D shape, for example, is a common sensible. So is the time at which an event occurs: it is possible to see, feel, and hear the moment at which a door knocker strikes the metal plate on a door. The existence of common sensibles, so understood, poses a prima facie challenge to the local distinctiveness thesis.

O'Callaghan (2014b, 2015) outlines a number of possible responses to the argument from common sensibles. One straightforward move is to drop to local distinctiveness for regional distinctiveness. Another way of responding appeals to modality-specific modes of presentation (Lopes 2000, Hopkins 2005, and Kulvicki 2007; for a skeptical assessment, see Bayne, 2014). Yet a third possible response appeals to modality-inflected phenomenal character, that is, phenomenal character that is partly determined by a "way" or "manner" of representation attaching to the modality itself (Chalmers, 2004). On this approach, perceptual modalities are introspectively different manners of relating to contents, and this is reflected in our experience of common sensibles across different modalities.

2.2 | The argument from optimizing multisensory integration

A distinct type of argument against the composite snapshot view targets the decomposition thesis. Optimizing multisensory integration (or O-integration for short) occurs when the initial estimates of a property provided by different modalities are weighted by their relative reliability and combined in a way that optimizes, that is, reduces the variance

in, the final perceptual estimate of that property (Ernst, 2012 provides a helpful overview). Examples of O-integration include the ventriloquism effect (Bertelson & de Gelder, 2004) and the sound-induced flash illusion (Shams, Kamitani, & Shimojo, 2002). For discussion, see Section 2.2 of the first part of this entry.

Spence and Bayne, 2014 distinguish between *causal* and *constitutive* conceptions of O-integration (see also Macpherson 2011: 446–449). According to the (purely) causal conception, the conscious end products of O-integration are wholly unisensory. According to the constitutive conception, by contrast, O-integration is reflected in conscious perceptual content that cannot be decomposed into distinct, unisensory layers. For example, the result of O-integration in the ventriloquism effect cannot always be decomposed into a merely visual plus a merely auditory experience of an event's location. Instead, O-integration may result in an inherently *audio–visual* representation of the event.

A circumspect assessment of relevant phenomenological and experimental evidence, a number of theorists have argued, is consistent with the purely causal view of O-integration (Deroy et al., 2014; Nudds, 2014; O'Callaghan, forthcoming; Spence & Bayne, 2014). Nudds, 2014, for example, writes: "this kind of integration does not undermine the idea that perceptual states are modality specific: that although the representational contents of the perceptual states of one modality are influenced by the contents of the perceptual states of another modality, our perceptual awareness of distal objects is explained in terms of modality-specific object representations" (183). And here is O'Callaghan, forthcoming: "coordinated perceptual awareness across the senses… is compatible with multisensory perceptual awareness being exhausted by that which is associated with each of the respective modalities along with whatever accrues thanks to mere co-consciousness" (manuscript draft: 8–9). If this is right, then the ventriloquism effect and other cases of O-integration do not equip us with compelling evidence against the decomposition thesis.

2.3 | The argument from intermodal feature binding

An alternative argument against the composite snapshot view is based on the possibility of intermodal binding awareness. There has been a large amount of research on intramodal feature binding, for example, on the attribution of shapes, colors, sizes, textures, and other features to objects by the visual system (Robertson, 2003; Treisman, 1999, 2003). That perceptual experiences within each of the modalities exhibit feature binding is relatively uncontroversial. There is substantial debate, however, about whether features experienced through different modalities can perceptually appear to be bound to the same object. If such intermodal binding awareness occurs, then "it is possible multimodally to perceptually experience the *apparent co-instantiation* of attributes perceived through different senses" (O'Callaghan, 2015: 557, emphasis added). Skeptics about intermodal binding awareness include Fulkerson, 2013, Connolly, 2014, Deroy et al., 2014, Spence & Bayne, 2014, and Nudds, 2016.

A first, empirically motivated argument for intermodal binding awareness appeals to research on multisensory "object files." An object file is a short-lived, perceptual representation that functions to keep track of an object across time and space and to store information about its potentially changing properties (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992, Noles, Scholl, & Mitroff, 2005). One source of evidence for the existence of such representations comes from studies of object-specific preview effects:

> *In a typical object reviewing display, a small number of objects (small outlined boxes) are initially presented, and letters are then displayed within them. The letters then disappear and the objects briefly move about the screen. When they halt, a final letter is displayed within one of the objects, and the observer's task is simply to name that letter as quickly as possible. This response is typically slightly faster when the letter matches one of the initially presented letters.... However, observers are faster still to name the final letter when it is the same letter that initially appeared on that object, as compared with when the final letter initially appeared on a different object—an* object-specific preview benefit *(Noles et al., 2005: 325).*

There is evidence that object-specific preview benefits can occur intermodally (Jordan, Clark, & Mitroff, 2010; Zmigrod & Hommel, 2011; Zmigrod, Spapé, Hommel, & B., 2009). As Bayne, 2014, O'Callaghan, 2015, and Spence

& Bayne, 2014 point out, however, the empirical evidence for the existence of multisensory object files is inconclusive with respect to the existence of non-modality-specific object representations in conscious experience. In an experiment by Mitroff, Scholl, and Wynn (2005), for example, subjects were shown an ambiguous bouncing/streaming display in which two moving objects could either be perceived to bounce off or stream through each other when their paths crossed. Mitroff and colleagues found that when the displays were designed so as to yield a strong bias toward the streaming interpretation, "there was nevertheless a strong OSPB in the opposite direction—such that the object files appeared to have 'bounced' even though the percept 'streamed'" (67). They interpreted this result as showing that the contents of conscious visual experience need not directly correspond to the information stored in visual object files. More relevantly, Zmigrod and Hommel (2011) investigated the relationship between audio–visual feature binding and conscious perception of audible and visual features as belonging to the same event. They conclude that "binding effects were entirely unrelated to conscious perception and did not even decrease in size when the bound features were perceived as separate events" (592).

Current experimental work on intermodal feature binding, then, does not provide uncontroversial support for intermodal binding awareness. For this reason, Bayne (2014) proposes to focus instead on what he takes to be one of the phenomenologically salient effects of optimizing multisensory integration. His example involves the sound-induced flash illusion:

> *…it seems to me that the perceptual experience that one has in the context of the sound-induced flash illusion does not leave it as an open question whether the flash and the beep are manifestations of a single event, but instead imposes this requirement on one's environment..... claims about the numerical identity of perceptual objects are built into the content of one's perceptual experience in the sound-induced flash illusion and many other examples of [multisensory integration] (Bayne, 2014: 26–27).*

Kubovy and Schutz (2010) offer a similar assessment:

> *To use phenomenological evidence [for audio-visual binding awareness] we would need a clear criterion for saying when an acoustic event and a visual event were bound to form a single audio-visual object. Some cases are clear. The sound and sight of a glass shattering leave no doubt in our minds that what we heard and saw was caused by the same physical event (54).*

The problem, as often is the case in phenomenological disputes, is that other theorists simply deny the existence of any such clear, introspective evidence. Spence and Bayne (2014), for example, write:

> *we think it is debatable whether the 'unity of the event' really is internal to one's experience in these cases, or whether it involves a certain amount of post-perceptual processing (or inference). In other words, it seems to us to be an open question whether, in these situations, one's experience is of a MPO [multisensory perceptual object] or whether it is instead structured in terms of multiple instances of unimodal perceptual objects (119).*

One phenomenologically-motivated way of attempting to move forward here, developed by O'Callaghan, utilizes the *method of phenomenal contrast*: "Take a pair of cases that controls for [experienced] spatio-temporal features and for other aspects of perceptual phenomenology. A case in which you 'get' the perceptual effect of intermodal binding awareness may contrast in character with an otherwise similar one in which you do not" (O'Callaghan, 2014a: 86). One example, he suggests, is the experience of ventriloquism:

> *You may seem to hear the visible puppet speaking, even if you are not taken in. Contrast this with a poor attempt at ventriloquism, in which it is perceptually evident that the visible puppet is not what you hear (O'Callaghan, 2014a: 83).*

Another example involves the experience of a watching a movie when the timing of the soundtrack is off:

> *The* alignment *matters. The dramatic phenomenological difference between the perfect soundtrack and the very poorly aligned soundtrack stems in part from perceiving audible and visible features as belonging to something common in the coincident case but not in the misaligned case (O'Callaghan, 2014a: 85).*

The skeptic about intermodal binding awareness, however, will regard O'Callaghan's interpretation of these cases as question begging. It is equally plausible, she will insist, that the experience of successful ventriloquism can be decomposed into visual and auditory experiences that are spatially and/or temporally coordinated with one another. In other words, what is absent when ventriloquism is unsuccessful is not intermodal binding awareness, but rather a sufficient degree of *correspondence* between visually and auditorily experienced (i.e., perceptually apparent), spatiotemporal features. Controlling for differences in such features should eliminate any phenomenal contrast between cases of successful and unsuccessful ventriloquism.

O'Callaghan is prepared for this skeptical response. "Intermodal binding awareness," he writes, "may depend not just on spatio-temporal cues, but also on factors such as whether and how the subject is attending, the plausibility of the combination or the compellingness of the match, and whether the subject expects one event or multiple events to occur.... *fixing spatio-temporal features does not by itself suffice in context to fix whether intermodal binding occurs*" (2014a: 85, emphasis added). In support of this conclusion, he points to an influential study of temporal ventriloquism by Vatakis and Spence, 2007. In temporal ventriloquism, "visual stimuli are 'pulled' into approximate temporal alignment with the corresponding auditory stimuli" (Vatakis & Spence, 2007: 744). What Vatakis and Spence found is that subjects found it easier to make temporal order judgments (TOJs) when auditory and visual speech stimuli were gender-mismatched, for example, a female face presented with a male voice, than when they were gender-matched, for example, a male face presented with a male voice. They interpreted this result in terms of a high-level "unity effect" (Welch & Warren, 1980; Warren 1999) on audio–visual multisensory integration: because male faces and male voices are categorized as belonging together, subjects' visual experience of the time at which the face's lips begin to move is pulled into alignment with their auditory experience of the time at which the voice is heard.

Contrary to O'Callaghan, however, Vatakis and Spence's study does not motivate the view that intermodal binding awareness can vary across cases in which experienced, spatiotemporal features are held constant. Although it is correct that differences in subjects' TOJs across conditions were not determined by differences in spatiotemporal cues (Vatakis and Spence "attempted to minimize any such bottom-up differences in the integration of the auditory and visual speech stimuli in the present study by carefully matching the [objective] timing of the visual and auditory events used to make the matched and mismatched videos" (2007: 753)), the gender-related unity effects discovered in the experiment were clearly on subjects' awareness of intermodal temporal relations.

In support of intermodal binding awareness, O'Callaghan (2014a, forthcoming) also points to studies that suggest that different types of O-integration can be selectively disrupted. First, subjects with autism spectrum disorder (ASD) do not use visual information to disambiguate audible speech in a neurotypical way. Among other things, children with ASD exhibit a much weaker McGurk effect than children without ASD (Mongillo et al., 2008). One explanation of deficits in audiovisual speech processing in ASD, however, is that subjects with ASD are significantly impaired in estimating the relative timing of auditory and visual speech signals (Brock et al., 2002; Stevenson et al., 2014). The same point can be made in connection with another example involving a patient, AWF, who cannot integrate auditory and visual speech information, described by Hamilton et al., 2006. The problem with this example is that AWF's linguistic deficit is explicitly characterized as a "perceived temporal asynchrony between vision and audition" (Hamilton et al., 2006: 71). If this is correct, then the relevant studies do not present persuasive empirical evidence for the absence of intermodal binding awareness in the presence of perceived, intermodal, spatiotemporal congruence.[2]

2.4 | The argument from novel feature instances

Some features may have instances that can be perceived only by means of one sense. Color, for example, may be perceptible only by means of vision. If so, then color is a *proper sensible.* Other features may have instances that can be

perceived by means of two or more senses. Shape, for example, can be perceived by means of both vision and touch. Shape, for this reason, is traditionally classified as a *common sensible*. Yet other features may have instances that can only be perceived using multiple senses in concert, instances that are "accessible only multisensorily" (O'Callaghan, forthcoming: 18). If so, then the composition thesis is false.

Spatial relations between objects experienced using different modalities are sometimes cited as an example. Tim Bayne (2014) writes: "one can be aware of the sound of a siren as being to the left of a visually presented dog. What... is the phenomenal character of one's awareness of this spatial relation? Clearly it could be neither purely visual nor purely auditory" (2014: 20). As O'Callaghan 2014 points out, however, this sort of case may just involve experiencing spatial locations or directions in different modalities co-consciously, rather than experiencing genuinely novel intermodal spatial relations.

A more compelling case involves evidence for intermodal meter perception. A recent study by Huang et al. (2012) found that auditory and tactile sequences were coherently grouped by musically trained subjects performing a meter recognition task. Importantly, neither channel by itself produced a coherent meter percept: in other words, the meter percept generated by audio-tactile grouping was novel relative to the intramodal sequences considered in isolation. Such intermodal meter perception, Charles Spence writes, "constitutes one of the first genuinely intersensory Gestalten to have been documented to date" (Spence, 2015: 647).[3] O'Callaghan, 2015 makes a convincing case that other types of features may have instances that are accessible only multisensorily. The examples that he discusses include intermodal causal relations and intermodal apparent motion.

2.5 | The argument from novel feature types

As O'Callaghan points out, the argument against the composite snapshot conception from novel feature instances is limited in the following way:

> *You can perceive spatial, temporal, and causal relations through vision, touch, or hearing alone. Since these feature types are familiar from unisensory contexts, perceptual awareness of their intermodal instances need not be multisensory in a deeper respect....*
>
> *...The arguments above demonstrate that... not every multisensory episode is just the co-conscious sum of its modality-specific parts. However, they do not show that it is not possible to account for multisensory perceptual awareness, even of novel feature instances, just in terms of (unimodal or amodal) features that unimodal perceptual experiences could have. And so, we might still say that the qualitative components of phenomenological character are not in this respect deeply multisensory (O'Callaghan, forthcoming: 30).*

A much stronger argument would undertake to show that there are types of features that can only be perceived multisensorily. Flavor properties, on one account (Smith, 2015; O'Callaghan 2014, 2015), answer to this description. According to the account, a substance's flavor is not represented by any single modality functioning in isolation. Instead, it depends on the combination of inputs from taste and retronasal olfaction; thermal and somatosensory cues; as well as information concerning chemical irritation and nociception supplied by the trigeminal system (Auvray & Spence, 2008).[4]

To use terminology introduced in the first part of this survey article, flavor perception may be an example of non-optimizing, "generative" multisensory integration (or G-integration for short). In G-integration, combining sources of environmental or bodily information from different modalities gives rise to the representation of a genuinely novel and "deeply multisensory" type of feature-one that could not be represented by any of the contributing modalities functioning in isolation.

Another possible example of G-integration, I suggested, is the experience of location (egocentric distance and direction) in external space. When G-integrated with sources of proprioceptive information, visual, auditory, and

tactile signals are all capable of representing the locations of objects in the distal environment using common, body-relative spatial reference frames (Briscoe, 2008, 2009; Briscoe & Schwenkler, 2015; Clark, 2011; Matthen, 2014, 2017). This is important not only because it supports our experience of a stable spatially structured world accessible through different senses, but also because it enables the crossmodal cuing of selective spatial attention (Spence, 2010; Spence & Ho, 2015) and the registration of crossmodal spatial congruence necessary for certain cases of O-integration (for further discussion, see the first part of this entry).

3 | OTHER PHILOSOPHICAL ISSUES

The last section of this review critically examined a number of arguments against the composite snapshot conception of perceptual experience. This section focuses on the relevance of research on multisensory processing for debates about the modularity of perception, cognitive penetration, and the individuation of the senses.

3.1 | Multisensory processing and modularity

Jerry Fodor (1983, 2001) influentially argued that perceptual "input analyzers" are modular, informationally encapsulated cognitive mechanisms. Vision, for example, functions to form representations of distal layout solely on the basis of sources of information in its proprietary database (think here of the different types of learned or innate "prior knowledge" to which Bayesian vision scientists make appeal) combined with afferent outputs from retinal transducers as well as efferent outputs from the oculomotor system.

At first blush, the different types of multisensory processing surveyed in the first part of this review challenge the Fodorian thesis that the senses are modular. The challenge, however, may be more apparent than real. Arguably, the key distinction for Fodor is between systems that, in principle, have unlimited access to stored, conceptual and non-conceptual information and systems that, by contrast, operate on a *restricted* class of inputs: "the claim that input systems are informationally encapsulated is equivalent to the claim that the data that can bear on the confirmation of perceptual hypotheses includes, in the general case, *considerably less* than the organism may know.... [Input systems do] not have access to *all* of the information that the organism internally represents" (Fodor, 1983: 69, emphasis added; see Burnston & Cohen, 2015 for an insightful discussion). Unlike systems supporting, for example, personal-level practical reasoning, input-analyzing systems typically do not have computational access to the contents of the subject's beliefs, intentions, and desires. They are cognitively impenetrable in the sense of Pylyshyn, 1999. So, even if there are "horizontal" information-sharing links between the senses at multiple computational levels (Shams & Kim, 2010), perceptual processing may be largely insensitive to much of the conceptually structured information available to central systems involved in high-level, theoretical inference and practical decision making (Deroy, 2015).

3.2 | The unity assumption and cognitive penetration

Evidence of certain high-level "unity effects" on optimizing multisensory integration (O-integration) complicate this picture, however. Some relevant experimental findings include the following:

A. A subject's belief that an auditory and visual stimulus "belong together," it has been suggested, can enhance audio–visual O-integration (for reviews, see Welch & Warren, 1980; Bertelson, 1999; and Bertelson & de Gelder, 2004). Visual biasing of auditory localization (ventriloquism), as evidenced by a pointing task, for example, has been reported to be greater when a kettle visibly emitting steam is paired with a whistling sound than when lights and meaningless tones are paired (Jackson, 1953).

B. O-integration of visual and haptic shape estimates can occur when subjects indirectly see their hand touching an object in a mirror. In the relevant experimental condition, a mirror was used to introduce an apparent separation of about 16 cm between the felt and the visible location of the object (Helbig & Ernst, 2007). The experimenters

interpret this result as showing that "knowledge that what we see is what we feel is sufficient for integration to occur, even in the case of spatial discrepancy" (1524).

C. Temporal ventriloquism occurs when auditory and visual speech stimuli are gender-matched, for example, a male face presented with a male voice, but not when they are gender-mismatched, for example, a female face presented with a male voice (Vatakis & Spence, 2007). Because male faces and male voices are categorized as "belonging together," subjects' visual experience of the time at which the face's lips begin to move is seemingly pulled into alignment with their auditory experience of the time at which the voice is heard (see Section 2.3 above).

Collectively, these results have been taken as evidence that cognitive states like beliefs can "penetrate" perceptual experience by influencing the O-integration process. There is room for skepticism, however. With regard to A, a number of studies have failed to find that familiarity or other cognitive factors influence the magnitude of the ventriloquism effect (for reviews, see Bertelson, 1999 and Bertelson & de Gelder, 2004). Contradictory findings may be explained, in part, by the fact that realistic pairings (e.g., kettles and whistles) are informationally rich in comparison with artificial pairings (e.g., briefly flashed lights and tones). In particular, they "have a greater internal temporal coherence and temporally varying structure" (Vatakis & Spence, 2007: 745; see also Deroy, 2014, 2015). It is also possible that findings suggestive of cognitive penetration actually reflect effects of background knowledge or experimental demands on post-perceptual judgments or motor responses (Bertelson 1994, 1999; Welch, 1999; Vatakis & Spence, 2007; Firestone & Scholl, 2015).

Turning to B, Helbig and Ernst suggest that "because participants were familiar with mirrors and because they saw their fingers exploring the object in the mirror, they had good reason to *believe* that both sensory inputs originated from the same physical object (2007: 1526, emphasis added). This may be so, but lower-level, perceptual factors may have been more directly responsible for the O-integration of visual and haptic size estimates in their experiment. In particular, even if visual and haptic signals were spatially discrepant, they exhibited a high degree of temporally coordinated structure. Further, it is not obvious that in order to solve the multisensory causal inference problem here ("Is the object visible in the mirror identical to the object being touched?"), the perceptual system must rely on subjects' beliefs about the etiology of mirror images. "What is required in using a mirror," Ruth Millikan suggests, "is only that one accommodate governance of one's perceptions and guided motions to a new semantic mapping function in taking account of the relation of seen objects to oneself. In the rearview mirror, I directly see that there is a car behind me. The car behind guides my motion in relation to it appropriately and directly" (Millikan, 2004: 122–123; see Schwitzgebel, 2014 for a similar assessment).

Two points are important. First, sophisticated conceptual representations such as beliefs, on this account, are not required for successful mirror use. Consistent with this conclusion, there is evidence that a number of nonhuman animals, including chimpanzees, marmosets, elephants, and magpies, are capable of using mirrors instrumentally to solve certain problems (for a review, see Gieling et al. 2014). Second, for subjects familiar with mirrors, objects seen in a mirror may look to be located where they really are. When a subject sees a car in her rearview mirror, it visually appears behind her, not in front of her. Similarly, when a subject sees her hand in a mirror, it may visually appear to be located where she proprioceptively represents it as being located. But, if this right, then it would be incorrect to interpret Helbig and Ernst's experimental findings as evidence that "knowledge that what we see is what we feel is sufficient for integration to occur, even in the case of spatial discrepancy."[5]

C, in my view, provides the most compelling evidence that O-integration can be influenced by high-level, cognitive or "non-structural" factors. Subsequent studies by Vatakis and colleagues strongly suggest, however, that audio–visual speech processing may be special. Vatakis and Spence (2008) found no unity effect on TOJs for pairs of auditory and visual nonspeech stimuli that either matched (e.g., the sight of a key being struck on a piano heard with the appropriate sound) or mismatched (e.g., the sight of a hammer smashing a block of ice dubbed with the sound of a bouncing ball). The unity assumption, they write "does not seem to influence people's temporal perception of realistic, multisensory, non-speech stimuli" (19). Relatedly, Vatakis et al. 2008 report that the unity effect does not influence audio-

visual O-integration when auditory stimuli are monkey vocalizations or human vocalizations of nonspeech sounds. For present purposes, the main point is that the role apparently played by cognitive factors in audio-visual speech processing does not necessarily generalize to other cases of O-integration.

3.3 | Individuating the senses

Research on multisensory processing and multisensory consciousness complicates philosophical attempts to individuate the senses, that is, to explain what distinguishes one sense from another (for helpful overviews, see the essays collected in Macpherson, 2011b, Fulkerson, 2014, Stokes, Matthen, & Biggs, 2014, and Matthen, 2015). Grice, 1962 suggested four, nonmutually exclusive ways of approaching this problem:

1. *The Represented Properties Criterion:* each sense is individuated by the objects and/or the range of features (properties and relations) that it represents.
2. *The Phenomenal Character Criterion:* each sense is individuated by the phenomenal character of the experiences to which it gives rise.
3. *The Proximal Stimulus Criterion:* each sense is individuated by the range of proximal stimuli to which its transducer systems evolved or have learned to respond.
4. *The Sense Organ Criterion:* each sense is individuated by a distinct set of transducer systems and perceptual information-processing mechanisms.

According to a maximally skeptical view, pervasive interaction and information-sharing between brain regions involved in processing outputs from retinal ganglion cells, cutaneous mechanoreceptors, and other peripheral transducers renders the very idea of distinct senses incoherent. On this view, we need to move "beyond perceptual modality" (Shimojo et al., 2001): there is just one complex perceptual system responsive to many different types of proximal stimulus information and, correspondingly, just one "metamodal" experience of the world. The Gricean criteria neither separately nor collectively allow us to construct a theoretically productive taxonomy of the senses.

A number of philosophers have criticized this maximally skeptical view. Fiona Macpherson (2011a, 2014) argues that applying the four Gricean criteria in conjunction permits us to construct a fine-grained, multidimensional space of actual and possible senses, one that takes into account the diverse ways in which perceptual systems cooperate and share information with each other. Matthew Fulkerson (2014) makes a strong case for *sensory pluralism*, the view that there are multiple, theoretically productive ways of "carving up" the perceptual modalities and their interactions. On this view, we should not expect a single, unified account of each sense. "Relative to a purely physiological criterion," for example, "we can categorize touch and temperature as separate modalities, but when they function reliably to bring awareness of wetness and material composition they are better categorized as part of a single haptic system" (Fulkerson, 2014: 3). Mohan Matthen (2015) argues that the senses can be usefully individuated, in part, by the interconnected, activities of purposeful, environmental exploration that they involve. Full acknowledgment of the diverse forms of multisensory processing and their impacts on perceptual consciousness as well as the various investigative contexts in which research on perception is undertaken, if these philosophers are right, makes individuating the senses a far more complicated project than traditionally supposed, but, contrary to the maximally skeptical view described above, it does not render that project incoherent.

ACKNOWLEDGEMENTS

I am very grateful to Casey O'Callaghan for comments that improved this review article. I would also like to thank Richard King, Fiona Macpherson, Mohan Matthen, and Matthew Nudds for helpful discussions at *The Senses and Crossmodal Perception* conference held at the University of Berne in October 2016.

Endnotes

[1] I follow O'Callaghan in defining a "mere" experience of a modality *p* as one that allows the subject to have had earlier experiences in modalities other than *p* but requires that the subject's overall perceptual experience remains wholly of *p*, while the experience occurs. This way of formulating the decomposition thesis notably allows for the possibility that certain multimodal "unity" relations may contribute to the phenomenal character of conscious experience (Bayne & Chalmers 2003; Bayne 2010, 2014). For skepticism about this idea, however, see Bennett and Hill (2014).

[2] I should emphasize that O'Callaghan's aim is not decisively to refute skeptics about intermodal binding awareness (O'Callaghan, forthcoming). Rather his argument is that a number of independent sources of evidence converge in support of the existence of such awareness. My own assessment, as should be clear, is less optimistic.

[3] For brain-imaging evidence that significantly overlapping neural networks and representations support rhythm perception across audition, vision, and touch, see Araneda et al., forthcoming.

[4] Richard Stevenson (2014) writes: "There is widespread agreement that when naive participants consume food, their experience of flavor is of a perceptual whole, something more than just taste, smell, and somatosensation... McBurney (1986) has offered a related description, referring to flavor as a perceptual fusion, distinct from synthesis (complete loss of parts) and analysis (complete access to parts). *Fusion* refers to a percept in which the individual sensory components are not eclipsed but where something unique to that combination of components emerges" (489).

[5] But see Chalmers (in preparation) for an argument that nonillusory mirror perception involves the cognitive penetration of visual experience by beliefs.

WORKS CITED

Araneda, R., Renier, L., Ebner-Karestinos, D., Dricot, L., & De Volder, A. G. (forthcoming). Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream. *European Journal of Neuroscience*.

Auvray, M., & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and Cognition*, *17*(3), 1016–1031.

Bayne, T. (2014). The multisensory nature of perceptual consciousness. In D. Bennett, & C. Hill (Eds.), *Sensory integration and the unity of consciousness* (pp. 15–36). Cambridge, MA: MIT Press.

Bennett, D., & Hill, C. (2014). *Sensory integration and the unity of consciousness*. Cambridge, MA: MIT Press.

Bertelson, P. (1999). Ventriloquism: A case of crossmodal perceptual grouping. *Advances in Psychology*, *129*, 347–362.

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. *In Spence and Driver*, 141–177.

Briscoe, R. (2008). Another look at the two visual systems hypothesis: The argument from illusion studies. *Journal of Consciousness Studies*, *15*(8), 35–62.

Briscoe, R. (2009). Egocentric spatial representation in action and perception. *Philosophy and Phenomenological Research*, *79*(2), 423–460.

Briscoe, R., & Schwenkler, J. (2015). Conscious vision in action. *Cognitive Science*, *39*(7), 1435–1467.

Burnston, D., & Cohen, J. (2015). Perceptual integration, modularity, and cognitive penetration. In J. Zeimbekis, & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 123–143). Oxford: Oxford University Press.

Calvert, G., Spence, C., & Stein, B. (Eds) (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Chalmers, D. (2004). The Representational Character of Experience. In B. Leiter (Ed.), *The future for philosophy* (pp. 153–181). Oxford: Oxford University Press.

Chalmers, D. The virtual and the real. In preparation.

Clark, A. (2011). Cross modal links and selective attention. In F. Macpherson (Ed.), *The senses: Classic and contemporary philosophical perspectives*. Oxford: Oxford University Press.

Connolly, K. (2014). Making sense of multiple senses. In R. Brown (Ed.), *Consciousness inside and out: phenomenology, neuroscience, and the nature of experience* (pp. 351–364). Springer.

Deroy, O. (2014). The unity assumption and the many unities of consciousness. In D. Bennett, & C. Hill (Eds.), *Sensory integration and the unity of consciousness* (pp. 105–124). Cambridge, MA: MIT Press.

Deroy, O. (2015). Multisensory perception and cognitive penetration: The unity assumption, thirty years after. In J. Zeimbekis, & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 144–160). Oxford: Oxford University Press.

Deroy, O., Chen, Y., & Spence, C. (2014). Multisensory constraints on awareness. In *Philosophical transactions of the royal society B* (pp. 369–1641) 20130207.

Ernst, M. (2012). Optimal multisensory integration: Assumptions and limits. *In Stein*, 527–544.

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *20*(2015), 1–77.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (2001). *The mind doesn't work that way*. Cambridge, MA: MIT Press.

Fulkerson, M. (2013). *The first sense: A philosophical study of human touch*. Cambridge, MA: MIT Press.

Fulkerson, M. (2014). Rethinking the senses and their interactions: The case for sensory pluralism. *Frontiers in Psychology*, *5*, 1–14.

Ghazanfar, A., & Schroeder, C. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285.

Gieling, E., Mijdam, E., Josef van der Staay, F., & Nordquist, R. (2014). Lack of mirror use by pigs to locate food. *Applied Animal Behaviour Science*, *154*, 22–29.

Grice, P. (1962). Some remarks about the senses. In R. J. Butler (Ed.), *Analytic philosophy, first series*. Oxford: Basil Blackwell.

Helbig, H., & Ernst, M. (2007). Knowledge about a common source can promote visual-haptic integration. *Perception*, *36*(10), 1523–1534.

Jackson, C. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, *5*, 52–65.

Jordan, K. E., Clark, K., & Mitroff, S. R. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition*, *18*(4), 492–503.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219.

Kayser, C., & Shams, L. (2015). Multisensory causal inference in the brain. *PLoS Biology*, *13*(2), e1002075.

Körding, K., Beierholm, U., Ma, W., Quartz, S., Tenenbaum, J., & Shams, L. (2007). Causal inference in multisensory perception. *PloS One*, *2*(9), e943.

Kubovy, M., & Schutz, M. (2010). Audio-visual objects. *Review of Philosophy and Psychology*, *1*(1), 41–61.

Kupers, R., Pietrini, P., Ricciardi, E., & Ptito, M. (2011). The nature of consciousness in the visually deprived brain. *Frontiers in Psychology*, *2*(4).

Macpherson, F. (2011a). Taxonomising the senses. *Philosophical Studies*, *153*(1), 123–142.

Macpherson, F. (2011b). *The senses: classic and contemporary philosophical perspectives*. Oxford University Press.

Macpherson, F. (2011c). Cross-modal experiences. *Proceedings of the Aristotelian Society*, *111*(3), 429–468.

Macpherson, F. The space of sensory modalities. In Stokes *et al.* 2014, 432–461.

Matthen, M. Active perception and the representation of space. Stokes *et al.* 2014, 44–72.

Matthen, M. (2015). The individuation of the senses. In M. Matthen (Ed.), *The Oxford handbook of philosophy of perception* (pp. 567–586). Oxford: Oxford University Press.

Matthen, M. (2017). Is perceptual experience normally multimodal. In B. Nanay (Ed.), *Current controversies in philosophy of perception* (pp. 121–136). London: Routledge.

Millikan, R. (2004). *Varieties of Meaning*. Cambridge, MA: MIT Press.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition*, *96*(1), 67–92.

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of autism and developmental disorders*, *38*(7), 1349–1358.

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*(9), 488–496.

Noles, N., Scholl, B., & Mitroff, S. (2005). The persistence of object file representations. *Perception & Psychophysics*, *67*(2), 324–334.

Nudds, M. Is audio-visual perception "amodal" or "crossmodal"? In Stokes *et al.* 2014, 166–190.

Nudds, M. Cross-modal object perception. Lecture delivered at *The Senses and Crossmodal Perception: Aristotelian and Contemporary Perspectives Conference*, University of Bern, October 2016.

O'Callaghan, C. (2008). Seeing what you hear: Cross-modal illusions and perception. *Philosophical Issues*, *18*(1), 316–338.

O'Callaghan, C. (2014a). Intermodal binding awareness. In D. Bennett, & C. Hill (Eds.), *Sensory integration and the unity of consciousness* (pp. 73–103). Cambridge, MA: MIT Press.

O'Callaghan, C. (2014b). Not all perceptual experience is modality specific. In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and its modalities* (pp. 133–165). Oxford: Oxford University Press.

O'Callaghan, C. (2015). The multisensory character of perception. *The Journal of Philosophy*, *112*(10), 551–569.

O'Callaghan, C. (forthcoming). Grades of multisensory awareness. *Mind & Language*.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341–365.

Robertson, L. (2003). Binding, spatial attention and perceptual awareness. *Nature Reviews Neuroscience*, *4*(2), 93–102.

Rohde, M., van Dam, L., & Ernst, M. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research*, *29*(4–5), 279–317.

Schwitzgebel, E. (2014). The problem of known illusion and the resemblance of experience to reality. *Philosophy of Science*, *81*(5), 954–960.

Shams, L. (2012). Early integration and bayesian causal inference in multisensory perception. In M. Murray, & M. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 217–231). Boca Raton, FL: CRC Press.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*(1), 147–152.

Shams, L., & Kim, R. (2010). Bayesian priors and multisensory integration at multiple levels of visual processing: Reply to comments on "crossmodal influences on visual perception". *Physics of Life Reviews*, *7*(3), 295–298.

Shimojo, S., Scheier, C., Nijhawan, R., Shams, L., Kamitani, Y., & Watanabe, K. (2001). Beyond perceptual modality: Auditory effects on visual perception. *Acoustical Science and Technology*, *22*(2), 61–67.

Smith, B. (2015). The chemical senses. In M. Matthen (Ed.), *The Oxford handbook of the philosophy of perception* (pp. 314–352). Oxford: Oxford University Press.

Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, *1191*, 182–200.

Spence, C. (2015). Cross-modal perceptual organization. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 649–664).

Spence, C., & Bayne, T. (2014). Is consciousness multisensory? In D. Stokes, S. Biggs, & M. Matthen (Eds.), *Perception and its modalities* (pp. 95–132). Oxford: Oxford University Press.

Spence, C., & Driver, J. (Eds) (2004). *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press.

Spence, C., & Ho, C. (2015). Crossmodal attention: From the laboratory to the real world (and back again). In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 119–138). Cambridge, MA: MIT Press.

Stein, B. (Ed) (2012). *The new handbook of multisensory processing*. Cambridge: MIT Press.

Stein, B., & Meredith, M. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stokes, D., Matthen, M., & Biggs, S. (2014). *Perception and its modalities*. Oxford: Oxford University Press.

Stratton, G. M. (1897). Vision without inversion of the retinal image. *Psychological Review*, *4/5*, 341-360, 463-481.

Stratton, G. M. (1899). The spatial harmony of touch and sight. *Mind*, *8*, 492–505.

Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron*, *24*(1), 105–125.

Treisman, A. (2003). Consciousness and perceptual binding. In A. Cleeremans (Ed.), *The unity of consciousness: binding, integration, dissociation* (pp. 95–113). Oxford England: Oxford University Press.

Trommershäuser, J., Kording, K., & Landy, M. (Eds) (2011). *Sensory cue integration*. Oxford: Oxford University Press.

Vatakis, A., Ghazanfar, A., & Spence, C. (2008). Facilitation of multisensory integration by the "Unity effect" reveals that speech is special. *Journal of Vision*, *8*(9), 1–14.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744–756.

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the "unity assumption" on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*(1), 12–23.

Welch, R. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. *Advances in Psychology*, *129*, 371–387.

Welch, R., & Warren, D. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667.

Zmigrod, S., & Hommel, B. (2011). The relationship between feature binding and consciousness: Evidence from asynchronous multi-modal stimuli. *Consciousness and Cognition*, *20*(3), 586–593.

Zmigrod, S., Spapé, M., Hommel, B., & B. (2009). Intermodal event files: Integrating features across vision, audition, taction, and action. *Psychological Research*, *73*(5), 674–684.

AUTHOR BIOGRAPHY

Robert Briscoe is associate professor of Philosophy at Ohio University. His research focuses on topics in the philosophy of cognitive science as well as the philosophy of perception. Special interests include the role of action in perception, spatial representation, mental imagery, depiction, and pictorial experience. Prior to his appointment at Ohio University, he was a research associate at the MIT Initiative for Technology and Self and taught in the philosophy department at Loyola University in New Orleans. He has a BA from Columbia University and a PhD from Boston University.

How to cite this article: Briscoe RE. Multisensory processing and perceptual consciousness: Part II. *Philosophy Compass*. 2017;e12423. https://doi.org/10.1111/phc3.12423